Genetics of human cardiovascular traits
Verweij, Niek
1. Introduction
Cardiovascular disease is a leading health care challenge and the number one cause of death globally\(^1\). Cardiovascular disorders include diseases of the vasculature, the myocardium, the cardiac conduction system, and congenital heart disease. Considerable effort has been made in the past 50 years to define, identify and modify risk factors for cardiovascular disease, and to develop effective treatment. Heart failure is currently considered one of the most challenging cardiovascular diseases, with a prevalence of 1-2%. The impact of heart failure is expected to increase substantially with the aging of the population, and despite advances in pharmacological and device treatment of heart failure the prognosis remains very poor, 17%-45% of patients die within 1 year of hospital admission\(^2\). Limitations of our knowledge on the pathophysiology of cardiomyocyte dysfunction and heart failure susceptibility is a major obstacle to progress our understanding of factors increasing the risk for heart failure development and also to design innovative treatment strategies.

In contrast to relative stagnation of heart failure treatment, remarkable progress has been made in human genetics, enabled by the first draft of the human genome in the year 2000 and novel technologies for DNA analysis beyond the sequence itself. This, in combination with the collection of exceedingly large cohort studies and international collaboration has revolutionized the field of genetic research. From 2006 onwards, researchers have been performing genome-wide association (GWA) studies to uncover the genetic basis and increase our understanding of many different complex diseases and traits in humans. GWA studies typically involves several hundreds of thousands (up to millions) genetic variants across the genome that are each tested for their association with a disease or trait in large population studies. By massively scanning the genome for associations, which requires no pre-defined biological assumptions about potential causal genetic regions or genes, it is possible to uncover potentially new biology, otherwise impossible by conventional 'candidate gene' approaches. GWA studies of complex diseases and traits have recapitulated many of the known genes and biological pathways relevant for various diseases and traits and often identified many more unknown ones to be involved. The identification and relevancy of biological pathways becomes even greater as the number of associated loci increases\(^3\). Genetic regions identified in GWA studies were also able to highlight genes that are targets of established drugs, such as thiazolidinediones and sulfonylurea (PPARG, type 2 diabetes), statins (HMGCR, lipids), Abatacept (CTLA4, rheumatoid arthritis), Ustekinumab (IL12B, Psoriasis), Ustekinumab (IL12B, Psoriasis) estrogen pathway and Denosumab (RANKL, bone mass)\(^4,6\) and potential new or repurpose drugs such as BCL11A inhibitors (BCL11A, sickle cell disease), Eculizumab (heart failure, Age-related macular de-
geration) and Tocilizumab (IL6R, Coronary artery disease) 4–6, indicating that new efforts with the aim to identify even more genetic loci for the trait(s) of interest are likely to be fruitful and may provide new clues for therapy and novel biological insights. This dissertation focuses mainly on the search for novel genetic determinants influencing cardiovascular traits and in particular those that are related to (left) ventricular function to increase our understanding of the underlying biology.

There are some considerations to take into account while performing GWA studies. GWA studies have been designed and powered to detect genetic associations through linkage disequilibrium (LD) between genetic variants (also named marker variants or marker alleles) and the unknown causal variants. To successfully identify novel genetic variants influencing the heart, it is important to consider that the statistical power to detect genetic variants is a function of the effect size, causal allele frequency, the correlation of the marker allele with the causal allele and sample size 7. The statistical power is further influenced by presence or absence of ancestry-based and phenotype-based heterogeneity 8. The properties of the unknown causal genetic variants like frequency, effect sizes cannot be changed, but the power of detection can theoretically be maximized by (1) accurately determine the marker (and potentially causal) alleles e.g. by a high density genotyping chip or whole genome sequencing; (2) subse-

quently impute ‘missing’ variants with higher density reference panels in order to increase the correlation of the marker alleles with the causal allele and possibly impute causal variants; (3) control for differences in population (sub)structure; (4) maximize the sample size; i.e. genotype and phenotype many individuals; and (5) accurately define the phenotype, further explained in the next paragraph.

Many of the above mentioned parameters are often restricted by available resources and costs, but can also be restricted to the frequency of a disease, in the case of a case/control study. Researchers are often also restricted to the individuals, phenotypes and measurements already available in cohort studies. Some studies that have difficulties defining a standardized phenotype, have been less successful in identifying true genetic associations. Classic examples that have difficulties identifying genetic variants are genetic studies of mental-health-related diseases 9 and behavioral traits 10. Similarly, GWA studies for cardiovascular diseases are complicated by the variation in disease classifications between cohort studies and the diverse pathogenesis, caused by an accumulation of risk factors underlying certain cardiovascular disorders such as stroke 11,12, sudden cardiac death 13 and heart failure 14. Collecting stricter phenotype definitions appeared to be slightly more successful in identifying genetic association that can be replicated in an independent population but are often limited by the small number of phenotyped and genotyped individuals: e.g. ventricular fibrillation in the presence of acute ischemia 15, dilated cardiomyopathy 16,17 and Brugada syndrome 18, identifying a hand full genetic variants.

Alternatively, the concept of endophenotypes has been suggested to face the problem of identifying genetic risk factors for neuropsychiatric diseases, plagued by ‘fuzzy’ definitions. This concept comprises intermediate measures unseen by the naked eye along the pathway between disease and distal genotype 19. Endophenotypes ideally (1) extends the phenotype beyond the clinical disease by identifying indices of genetic risk that can be measured in individuals without symptoms and (2) may be closer to the primary effects of susceptibility genes (3) or both. In support of this concept is the theory that complex disorders can be thought of as the extremes of quantitative traits 20. Hence, variation in clinical phenotypes are likely to reflect, at least in part, a combination of quantitative variation in genetic risk for the disorder and the effect of modifying variants that affect the expression of the disease rather than the risk for the disease itself 21. Therefore studying endophenotypes of diseases may provide other, and possibly more, insight into human biology, than studying the clinical disease phenotypes alone. Like neuropsychiatric diseases, heart failure is also a ‘fuzzy’ phenotype. The complexity and clinical heterogeneity of heart failure has challenged the understanding of phenotype-genotype relationships 22. So, studying intermediate, heart failure related, traits will be logical next step.

A successful study utilizing endophenotypes of cardiovascular disease is the genetic study of circulating concentrations of natriuretic peptides in 14,743 healthy subjects, identifying the NPPA-NPPB locus, among others, and also confers risk of hypertension 23. These peptides are known to regulate blood pressure and are disturbed in many cardiovascular disease settings. Another example is the genetic variant in the SMIM1 locus that we initially have discovered in a GWA study of red blood cell indices 24 and later showed to affect expression of the clinically important, but rare, Vel-negative blood type 25. Similarly, parameters on the electrocardiogram (ECG) have been studied to identify genetic variants associated with cardiac conduction such as RR (heart rate)-, PR-, QRS- and QT-duration. These studies have identified a number of genes already thought to play a role in these traits, such as ion channels (SCN5A, KCNH2, KCNQ1, KCNJ2, KCNE1) and cardiac transcription factors (TBX3, TBX5, HAND1, NKX2-5). ECG parameters, primarily of PR-interval were also able to capture various genetic variants associated with atrial fibrillation. Few of these variants were associated with sudden cardiac death (SCD), which could be due to a combination of the phenotypic heterogeneity underlying SCD, generally small effects of ECG variants on conduction and possibly opposing
pleiotropic effects (some variants prolong QRS but shorten QT duration)\textsuperscript{13}.

This dissertation has one principle aim: to identify novel genetic loci associated with cardiovascular traits related to the development of Heart Failure and left ventricular function, thereby increasing our knowledge of the genetics and biology underlying these cardiovascular traits.


